

バイオサイエンスデータベースセンターの今後のあり方について（提言）

平成 29 年 3 月

国立研究開発法人科学技術振興機構

バイオサイエンスデータベースセンター運営委員会

（概要）

バイオサイエンスデータベースセンター（NBDC）は、ライフサイエンス分野における我が国のデータベースの一元的な統合を目指し、2011（平成 23）年 4 月に発足した。以来、日本のライフサイエンス研究から産出されるデータやデータベースが最大限に活用されるように、また、研究終了後に死蔵されることの無いように、これらのデータやデータベースを統合・整理するための研究開発とサービス提供を行ってきた。一元的な統合は道半ばであるが、これまでの NBDC の活動は我が国のライフサイエンス研究の成果を統合し発展させるために欠くことのできない基盤となっている。

本提言は、我が国のライフサイエンス研究におけるデータ基盤整備とその中で NBDC が果たすべき役割に関して本委員会が行った議論を取りまとめたものである。研究の進展等による変化には柔軟に対応しつつ、我が国全体としての課題には関係する省庁や機関と連携して積極的に解決にあたるよう、提言する。

～ 今後 5 年をめどに NBDC の活動を拡大・強化して取り組むべき内容 ～

1. 支援対象を、公開済みデータだけでなく未公開データまで拡大する－【新たな対象】
 - ・我が国のライフサイエンス研究データを網羅的に把握・可視化し、大型プロジェクト研究のデータが関係者間でいち早く最大限に利活用されるように支援する
2. データベース構築者の視点から、利用者の視点に転換する－【新たな利用】
 - ・ヒト由来データの共有と利活用は喫緊の課題であり、ヒトとモデル動物の研究データをゲノムから表現型まで統合し、利用者の要請に応えられるデータ基盤を確立する（当面の連携先として、東北メディカル・メガバンク計画、臨床ゲノム情報統合データベース整備事業、ゲノム医療実現推進プラットフォーム事業、AMED 微生物叢関連プロジェクト群、COI STREAM 健康・医療データ連携事業、などが想定される）
3. 新たな知識やイノベーションを生み出すデータベースを構築する－【新たな価値】
 - ・世界に伍する研究開発が実現するように、大規模データ解析や人工知能を含む多様な分野の研究者と共に、データベースの戦略的統合と利活用に取り組む

～ NBDC が関係する府省や機関と共に解決に当たるべき課題 ～

4. データ整備・統合の継続性を保障する体制の整備－【研究基盤には継続性が不可欠】
 - ・恒久的な財政措置（国の研究開発予算のうち一定割合を確保するなど）
 - ・バイオインフォマティクス人材の育成とキャリアパス設計
5. 我が国の顔としての求心力・発信力の強化－【プレゼンス強化には一元化が不可欠】
 - ・継続性のある一元的な統合データベース機関の設立
 - ・我が国を代表する統合データベースとしての国際活動

I 現状認識

バイオサイエンスデータベースセンター (NBDC) は、2009 (平成 21) 年の「統合データベースタスクフォース報告書」(総合科学技術会議ライフサイエンス PT 統合データベースタスクフォース) を受けて、2011 (平成 23) 年 4 月に発足した。同報告書は我が国の「恒久的かつ一元的な統合データベース」の整備を提言しており、その実現に向けた中核的な活動を担う組織として、NBDC は設立されたものである。

NBDC はこうした設立の趣旨に精力的な活動で応え、当時問題とされていた課題の一部は解決された。省間連携体制を構築し、国内の公開データベースの所在情報を網羅するデータベースカタログやデータベース横断検索を実現した。また、個人情報保護しつつヒト由来データを共有化できる NBDC ヒトデータベースも構築した。さらに、データベース統合に必要な研究開発を実施してきた^{*}。一方で、困難な課題は依然として残されており、ライフサイエンス研究の進展などによって新たな課題も生じている。

^{*}例として、PDBj や KEGG MEDICUS 等へのファンディング、TogoGenome の開発、ゲノム情報の RDF 化モデルの考案、が挙げられる。詳細は (参考)「3. 事業の概要と成果」を参照。

～「統合データベースタスクフォース報告書」で指摘され、解決していない課題～

- 欧米と比較して、データベース統合に向けた取組みの規模や体制が不十分
- ・司令塔機能として提言された「統合データベース推進本部 (仮称)」が設置されていない。予算規模は、拡大するどころか縮小しつつ、ピーク時 (平成 19 年度) ^{*}の 4 割程度になっている。

^{*}NBDC の前身事業である「バイオインフォマティクス推進センター事業」と「文部科学省統合データベースプロジェクト」を合算。

- 恒久的な財政措置がなく、データの継続的な利用が保障されていない
- ・国際的に評価の高いデータベースであっても NBDC からのファンディングに依存。また、ファンディングでは代表研究者が退職した後の継続が保障されていない。

～ 新たな課題～

- 一層多くのデータが産出され、データの取扱いに抜本的な変化が必要になった
- ・データサイエンスの新たな展開が想定され、国を挙げた戦略的な取組みが必要。
- モデル生物だけでなく、ヒトに由来する研究データも盛んに産出されるようになった
- ・個人情報保護に配慮した取扱いが必要であることに加え、個人差や環境の違いなど複雑性が高く、データ統合の難度が高い。
- バイオインフォマティクス人材の育成が進んでいない
- ・予算が認可されなかったため、NBDC では本格的な取組みは実施できていない。また、雇用も不安定で、育成・輩出が進まない大きな原因となっている。

こうした課題には NBDC の活動を改善・強化することである程度対応できる部分もあるが、NBDC 単独では解決が難しい部分もある。そこで、NBDC と我が国全体の 2 つの観点から取り組むべき内容を提言として取りまとめた。

II 今後 5 年をめどに NBDC の活動を拡大・強化して取り組むべき内容

ライフサイエンス研究の進展や研究を取り巻く環境の変化によって、NBDC に求められる役割は急速に変化している。利用者からの要請に柔軟に対応するため、今後 5 年をめどに NBDC の活動を拡大・強化して取り組むべき内容を提言する。

1. 支援対象を、公開済みデータだけでなく未公開データまで拡大するー【新たな対象】

オープンサイエンスが世界的な潮流となる一方で、知的財産権の確保等のためデータを直ちに公開することが難しい場合も多くある。NBDC では、公的研究資金で作成されたデータベースの所在情報を「データベースカタログ」として整理し提供している。国内の公開データベースに関してはほぼ網羅している状況であるが、未公開データは対象としていない。データが非公開である段階から所在を把握し顕在化させることによって、データの重複取得を防ぐと共に共同研究によるデータ利活用を促進する効果が期待できる。また、公開前から共有・利活用に適した形でのデータの整備を NBDC が支援できれば、より幅広く効率的な共有・利活用と円滑なデータ公開が可能となる。

(取り組むべき内容)

- 省庁の枠を越えて大型プロジェクト研究と連携し、公的研究資金を活用して産出されるデータについて、公開前から所在を網羅的に把握する
- 把握した所在情報について、わかりやすく整理し公開する仕組みを構築する
- 国の大型プロジェクトのうち支援効果が大きいもの*については、公開前から公開・共有を見据えたデータ作成・データベース構築や限定された研究者間のデータ共有を支援する
 - ※①日本の強みや戦略目標等の施策動向 ②産業利用可能性 ③利用頻度、を総合的に勘案し、多様なテーマ・分野から NBDC 自身で主体的に選定を行う
- データの所在把握や支援の企画・推進を行う研究員を増員するほか、連携するプロジェクトとの人的交流を実施し、人材の育成にも貢献する

2. データベース構築者の視点から、利用者の視点に転換するー【新たな利用】

ヒトに由来する研究データの産出が近年急速に増加している。次世代シーケンサの普及等、計測技術が進展して大量のデータが安価・高速に産出されるようになり、ゲノム医療を目指した研究が加速した。ゲノム医療研究は世界各国で基幹プロジェクトとして推進されており、ヒト由来データの共有と利活用は国際的かつ喫緊の課題である。我が国に関しては 2015 (平成 27) 年 4 月に日本医療研究開発機構 (AMED) が設立され、健康医療分野の公的研究費事業が集約されたことも対応が必要な大きな変化と言える。我が国のデータベース統合における中核機関として、NBDC は国内の主要プロジェクトからの要請に応えられるデータ基盤を確立すべきである。

(取り組むべき内容)

- 我が国の主要な大型プロジェクト研究*から産出されるデータが確実に NBDC ヒトデー

データベースに収録されるように、また、最大限有効に活用されるように、AMED 等と共同して企画立案を行う

※例として、東北メディカル・メガバンク計画、臨床ゲノム情報統合データベース整備事業、ゲノム医療実現推進プラットフォーム事業、AMED 微生物叢関連プロジェクト群、革新的イノベーション創出プログラム (COI STREAM) 健康・医療データ連携事業などが考えられる

- これまで主としてオミクス別に統合されてきたデータベース同士を統合し、ヒトとモデル動物の様々なオミクスデータや研究用試料情報^{*}を統合的に利活用できるデータベース等を開発する

※例として、ナショナルバイオリソースプロジェクト (NBRP)、オーダーメイド医療の実現プログラム (BBJ)、東北メディカル・メガバンク計画などが考えられる

- これまで以上に利用者の意見を収集し、サービスの充実と利活用促進を図る
- 当面は研究データの統合化を進めることが妥当であるが、我が国として時機を逃さず研究データと診療・健診データを統合的に利活用できるように、医療データの取扱いに関する国内外の動向を調査し、関係府省・機関等への協力を行う

3. 新たな知識やイノベーションを生み出すデータベースを構築するー【新たな価値】

公的研究資金を活用して得られたデータは国の資産であり、死蔵されることのないように、また、最大限に活用されるように、データの産出だけでなく統合・利活用まで含めて戦略的に取り組む必要がある。「統合データベースタスクフォース報告書」で司令塔機能として提言された「統合データベース推進本部 (仮称)」が設置されない状況の中、NBDC には国のライフサイエンスデータ戦略を先導する役割が求められている。大規模データ解析や人工知能研究といった近年発達しつつある分野の研究者も含め、多様な分野の研究者と共に、我が国として必要なデータを幅広く利活用できるように整備する必要がある。また、データ駆動型の研究手法は我が国では十分に普及しているとは言い難く、統合したデータの利活用が進むよう、データベース統合の重要性や価値を発信していくことも重要である。

(取り組むべき内容)

- ライフサイエンス分野のデータベース利用者に加え、大規模データ解析や人工知能等、近年発達しつつある多様な分野の研究者と共に、データベース統合がもたらす新たな意味と価値を追求し、我が国のライフサイエンスデータ戦略の立案を先導する
- 大規模データ解析や人工知能等も含めた多様な分野の研究者と研究企画段階から密接に連携して潜在的ニーズを発掘し、利用者とともに統合データの利活用に取り組む
- データベース統合に向けた研究開発を継続し、我が国として国際的に主導的地位を得ることが可能な統合化前の新興領域についても時期を逃さず統合化を支援する
- データベースやサービスの利用状況をよりよく把握するとともに、今後重要性が増すと考えられる画像やテキストなどの非構造化データへの対応についても引き続き検討する

なお、これらの取組みにあたっては、ライフサイエンス研究や周辺技術の進展に沿った活動が行われるよう定期的にレビューと見直しを行い、選択と集中の考え方にに基づき、人的資源や予算状況も踏まえて推進すべき活動やその優先度等の再検討を行う必要がある。

III NBDC が関係する府省や機関と共に解決に当たるべき課題

データベース統合は、我が国のライフサイエンス研究を支える基盤として、将来に亘り継続的に実施されるべきである。NBDC の約 6 年の活動によって我が国のデータベース統合は大きく進展したが、NBDC 単独では解決が難しい以下のような課題も残されている。これらの課題は、我が国としての継続性のある一元的な統合データベース機関が無いことに因るところが大きい。NBDC でも解決に向けた検討を継続するべきであるが、国としてもより広範な関係者を交えて国家戦略としての検討を深めていただくことが必要と考える。

4. データ整備・統合の継続性を保障する体制の整備－【研究基盤には継続性が不可欠】

研究基盤には継続性が不可欠であり、そのためには予算と人員の確保が必要である。しかし、扱うデータの量や期待される役割が増大する一方で、NBDC の予算は毎年減少し続けている。また、国際的に評価の高いデータベースであっても NBDC からのファンディングに頼った運営になっている状況も散見され、継続的な運用が保障されていないうえに、NBDC の予算減に伴って運営が年々厳しくなっている点も問題である。国際的に評価の高いデータベースに関しては NBDC が運用を継続できるような仕組みが検討されているが、その実現に必要な予算や人員を確保できる見通しが立っていない。さらに、データ基盤の維持・発展にはパーマnent雇用のスタッフによる長期的な取組みが不可欠であり、データベースに関連する人材の多くが任期付で雇用されている現状についても、改善を図るべきである。

(NBDC が取り組むべき内容)

- 例えばライフサイエンス分野の公的研究開発資金のうち一定割合を公的データベースの運用・維持に確保するなど、データの継続的利用を保障する方策が実現するように、関係府省に働きかけを行う
- データベース関連人材が雇用や任期の心配なく活躍できるような仕組みづくりに向けて、関係府省・機関に協力を行う
- 文部科学省やAMED等が実施する人材育成事業^{*}に対して、教材データの提供や講師派遣など、実施可能な協力を行う

^{*}例として、「第4次産業革命に向けた人材育成総合イニシアチブ」などが考えられる。

5. 我が国の顔としての求心力・発信力の強化－【プレゼンス強化には一元化が不可欠】

新たなデータサイエンスの展開が想定され、世界的にもオープンサイエンスの潮流が進む中、我が国の顔としてのデータベース機関や国を挙げた戦略的な取組みが必要である。

しかし、我が国には、NBDCのほか、ライフサイエンス統合データベースセンター（DBCLS）、公的資金で運営されている国際レポジトリ（国立遺伝学研究所 DDBJ センター、大阪大学蛋白質研究所 PDBj など）といった複数のデータベース関連機関が存在している。相互に連携し協働しているものの、利用者を含め国内外の関係者には違いや役割分担が理解しにくい状況であり、データや人材を集める求心力を出せずにいる。また、欧米と比較して人的資源・予算が桁違いに少ないことや戦略の欠如もあって、欧米の一元的で継続的な統合データベース機関に伍する国際活動が困難な状況にある。

（NBDC が取り組むべき内容）

- 我が国としての継続性のある一元的な統合データベース機関の設立など、欧米と肩を並べて活動できる体制の整備に関して、関係府省・機関と共に検討する
- 我が国としての体系的な、データ産出から統合や利活用まで含めた戦略の策定に向けて、関係府省・機関と共に取り組む
- データの共有範囲の拡大や迅速な公開に着実な進展をもたらす仕組みづくりに関して、関係府省・機関への協力を行う
- 関係機関と一丸となって、データベースセンターの認知度を高める活動を行う
- 関係機関でより一体的な活動が可能となるよう、これまで実施してきた国際活動[※]や更なる取組みに尽力し、日本の代表としての国際活動体制を強化する

※BioHackathon の主催や GA4GH（Global Alliance for Genomics and Health）への参画など。

バイオサイエンスデータベースセンター (NBDC) 運営委員会 委員名簿

- 堀田 凱樹 公益財団法人井上科学振興財団 理事長
- 伊藤 隆司 国立大学法人九州大学大学院医学研究院 教授
- 大江 和彦 国立大学法人東京大学大学院医学系研究科 教授
- 加藤 規弘 国立研究開発法人国立国際医療研究センター研究所
遺伝子診断治療開発研究部 部長
- 金岡 昌治 株式会社住化分析センター 代表取締役社長
- 篠崎 一雄 国立研究開発法人理化学研究所
環境資源科学研究センター センター長
- 城石 俊彦 大学共同利用機関法人情報・システム研究機構
国立遺伝学研究所 副所長
- 末吉 互 潮見坂綜合法律事務所 弁護士
- 諏訪 牧子 学校法人青山学院 青山学院大学
理工学部化学・生命科学科 教授
- 樋口 知之 大学共同利用機関法人情報・システム研究機構
統計数理研究所 所長
- 松田 秀雄 国立大学法人大阪大学大学院情報科学研究科 教授

○ : 委員長

平成 29 年 3 月現在

(参考)

バイオサイエンスデータベースセンターの発足までの経緯、事業概要と成果

1. 発足までの経緯

1. 1. バイオインフォマティクス推進センター (BIRD)

2001 (平成 13) 年 4 月、科学技術振興機構 (JST) にバイオインフォマティクス推進センター (BIRD) が設立された。内閣府の前身である総理府に設置された、科学技術会議ライフサイエンス部会ゲノム科学委員会の「ゲノム情報科学におけるわが国の戦略について」(2000 (平成 12) 年) における、バイオインフォマティクス推進の提言を受けてのことである。

BIRD では、データベースの構築や高度化・標準化のための研究開発を 2 期に亘って実施するとともに、創造的な生物・情報知識融合型の研究開発をファンディングにより支援し、これを基盤として新たな生物科学が樹立されることを目指した。また、これらの研究開発やファンディングの成果として得られたデータベースやツールを公開・提供するとともに、2002 (平成 14) 年度からは、それまで他事業で実施されていた、HOWDY や JSNP 等のデータベースの運用や公開もバイオインフォマティクス推進センター事業に一本化し、これらの成果の公開を継続して行った。

主な成果として、大阪大学の PDBj (Protein Data Bank Japan)、京都大学の KEGG (Kyoto Encyclopedia of Gene and Genomes)、国立遺伝学研究所の DDBJ (DNA Data Bank of Japan) など世界的に定評のあるデータベースが挙げられる。

1. 2. 文部科学省「ライフサイエンス分野における統合データベースプロジェクト」

一方で、2000 (平成 12) 年当時の予想を超えて多くのデータベースが作られるようになったこと、構築されたデータベースが散在しており効率的な利活用のための基盤整備が必要になったこと、等から、2005 (平成 17) 年 8 月、文部科学省 科学技術・学術審議会 研究計画・評価分科会 ライフサイエンス委員会の下に、データベース整備戦略作業部会が設置された。そして翌 2006 (平成 18) 年 5 月、同作業部会の議論を経て取りまとめられた報告書「我が国におけるライフサイエンス分野のデータベース整備戦略のあり方について」を受け、文部科学省において、2006 (平成 18) 年度より 5 年間の時限付きプロジェクトとして「ライフサイエンス分野における統合データベースプロジェクト」が開始された。2007 (平成 19) 年 4 月には、中核拠点として「ライフサイエンス統合データベースセンター」(DBCLS) が大学共同利用機関法人情報・システム研究機構に設置された。

統合データベースプロジェクトでは、ライフサイエンス、情報処理技術、ライフサイエンスデータベースの 3 分野の専門家による研究運営委員会が我が国のデータベース整備戦略を立案するほか、データベースや電子化された文献等から必要とする情報を容易に検索し

て引き出せるようにするための技術とその情報基盤の開発、さらには、データベースの統合化を支援するポータルサイトの開発や、大学と連携して人材を育成するなどの活動を行った。

1. 3. NBDC

データベース統合化の取り組みは、当初、各省で独自に進められていたが、2006（平成18）年頃から内閣府の主導のもと、省庁連携によるオールジャパンの協力体制を構築して進められた。2009（平成21）年5月には、「統合データベースタスクフォース報告書」（総合科学技術会議 ライフサイエンス PT 統合データベースタスクフォース）の中で、DBCLSとBIRDの一体的な運用を、JSTに新たに設置する組織で行うことが提言された。これを受け、2011（平成23）年4月、JSTにNBDCが発足し、ライフサイエンスデータベース統合推進事業が開始された。

2. 事業の目的と推進体制

「統合データベースタスクフォース報告書」において、センターに求められる機能として、「・・・文部科学省の統合データベース事業でROISが中心に取り組んできたライフサイエンスデータベースの整備（統合・維持・運用）、JSTが取り組んできたライフサイエンス分野における基盤的データベースの高度化・維持・運用を一体化して引き継ぎ、更なる発展を目指していくとともに、利用者が、各省等におけるデータベースも含めて我が国全体に散在するデータベースを有効に使うことのできる運用を目指した一元的なネットワーク化の構築を主なものとする。」とされた。

上記を受け、NBDCは、我が国におけるライフサイエンス研究の成果が、広く研究者コミュニティに共有かつ活用されることにより、基礎研究や産業応用研究につながる研究開発を含むライフサイエンス研究全体が活性化されることを目的として活動を実施している。具体的には、後段に述べる戦略の立案、ポータルサイトの構築・運用、データベース統合化基盤技術の開発、バイオ関連データベース統合化の推進、を業務の4つの柱としている。

なお、上記タスクフォース報告書では、平成23～25年度を第一段階と定め、統合データベースのためのセンターをJSTの新たな組織として整備することを提言している。平成26年度以降の第二段階については第一段階で整備した成果を考慮しつつ、引き続き総合科学技術会議（当時）で検討していくことが必要であるとされていた。これを受けて内閣府、文部科学省で検討がなされ、第二段階についてもNBDCを中心とした現行の体制で事業を推進することが了承された。

3. 事業の概要と成果

3. 1. 戦略の立案

データベースの整備・統合化の戦略企画や、データや技術のコーディネート、データベース統合化ガイドラインの策定、国内外との連携構築などにより戦略の立案を行っている。

成果としては、オールジャパンでのデータベース統合への取り組みのため、総合科学技術会議ライフサイエンス PT に設けられた統合データベース推進タスクフォース会合、統合データベース意見交換会、その後はライフイノベーション戦略協議会の場で NBDC の活動状況を報告し、推進方策についての議論を踏まえ活動方針を定めつつ着実な運営を実施してきたことが挙げられる。また、文部科学省・厚生労働省・農林水産省・経済産業省との連携体制を構築し、次項に述べるデータベースカタログや横断検索といったサービスを通じて、各省のデータベースを有効に活用できるような環境を整備した。

加えて、我が国においてデータベースにおけるヒト由来データの取り扱いに関する統一的な指針が未だ整備されていない状況であったところ、NBDC 運営委員会に倫理分科会、ならびにデータ共有分科会を組織し、2013（平成 25）年 4 月に「NBDC ヒトデータ共有ガイドライン」「NBDC ヒトデータ取扱いセキュリティガイドライン」を公開するとともに、国立遺伝学研究所と連携して、2013（平成 25）年 10 月よりヒトに関する研究データの共有を促進するためのプラットフォームである NBDC ヒトデータベースの運用を開始した。文部科学省の「次世代がん研究シーズ戦略的育成プログラム」（P-DIRECT）や「オーダーメイド医療の実現プログラム」（BioBankJapan）、東北メディカル・メガバンク機構（ToMMo）を含む国内の多様な研究開発プロジェクトから、2016（平成 28）年 9 月末までに 100 件のデータ提供申請があり、公開待機と審査・確認中を除く 44 件を公開（制限公開を含む）した。また、Nature 姉妹誌である Scientific Data 誌の推奨レポジトリ一覧に、米国 dbGaP や欧州 EGA とともに NBDC ヒトデータベースが掲載され、研究者コミュニティに周知されている。このように、我が国で唯一の、機微情報を含むヒトデータを広く共有する仕組みを持ったデータベースとして、NBDC ヒトデータベースは国内外で認知度を高めてきており、こうした実績が AMED 「疾病克服に向けたゲノム医療実現化プロジェクト」との連携開始に繋がっている。

3. 2. ポータルサイトの構築・運用

NBDC ポータルサイトの構築とその運用により、各府省や研究機関、大学に散在しているデータベースについて、データベースの所在や概要を閲覧・検索できるようにする「データベースカタログ」、散在しているデータベースを特許や文献と合わせて一括検索できるようにする「データベース横断検索」、国内の生命科学系研究者が作成したデータベースを検索・ダウンロードできるようにする「データベースアーカイブ」などを提供している。

NBDC ポータルサイトは 2011（平成 23）年度に開設し、アクセス数は 2013（平成 25）年度以降、各年度約 400 万ページビューとなっている。データベースカタログについては、

1,581件、横断検索については596件のデータベースを対象としており（いずれも2016（平成28）年9月末時点）、国内で公開されているデータベースを網羅する目途が立った。データベースアーカイブは、119件を掲載している（2016（平成28）年9月末時点）。

同2011（平成23）年には、文部科学省・厚生労働省・農林水産省・経済産業省と合同で、「生命科学系データベース統合のための合同ポータルサイト」を開設した。4省連携の取組みに関する情報発信に加えて、本合同ポータルサイトからも、データベースカタログ、横断検索、データベースアーカイブを利用可能としている。

3. 3. データベース統合化基盤技術の開発

データベース統合化の実現に向けて、基盤となる技術の開発を実施している。具体的には、データベースのRDF化、オントロジー（用語・概念の対応関係・階層関係を規定した用語集）整備や自然言語処理、解析ワークフロー、大規模データ処理等の技術開発を行っている。

RDF化技術の開発やオントロジーの整備等により、データベース構築者やバイオインフォマティクスに再利用性の高いデータを提供し、他との接続性の高いデータベースの開発やその解析が効率的に行えるよう基盤を整えた。また、RDFによるゲノム情報の統合データベースTogoGenomeを開発したほか、健康医療分野におけるデータ共有・研究促進のための国際プロジェクトに参画し、NBDCヒトデータベースの一部を国際的なヒトゲノム変異の検索システムGA4GH Beacon Networkに登録した。加えて、データベース統合化の国際的標準化とその実装を目指すワークショップBioHackathonや、国内のデータベース開発担当者を集めたワークショップSPARQLthonを開催し、データベース統合化技術の開発や利用促進を図ると共に、国内外の研究者・技術者との人的ネットワークを構築した。こうしたネットワークも活用しながら、欧州バイオインフォマティクス研究所（EBI）がEnsemblのゲノム情報をRDF化するに先立ち、EBIと共同してDDBJやTogoGenomeと共通のRDFモデルを考案し、ゲノム情報のRDFモデルの国際標準化に向けて中核的な役割を果たした。

3. 4. バイオ関連データベース統合化の推進

バイオ関連データベースの統合化を実現するため、統合化推進プログラムのファンディングによって、分野ごとのデータベース統合化等を推進している。

2011（平成23）年度以降、のべ22課題を採択し、分野別やオミクス別に整備されたデータベースはNBDCポータルサイトからアクセス可能な形で公開され、ライフサイエンス分野の研究開発において活用されている。顕著な成果の例としては、タンパク質構造データベースPDBj（国際的なデータベースwwPDBの一翼を担い世界の約22%のタンパク質構造データを掲載）、国際糖鎖構造リポジトリGlyTouCan（日本発の国際的な糖鎖構造データリポジトリシステムを開発、糖鎖関連論文の標準化に関する団体からの支持）、プロテオーム統合データベースjPOST（プロテオームの国際コンソーシアムProteomeXchangeにアジ

ア地域初の加入)、表現型データベース (国際マウス表現型コンソーシアムが保有するマウスの網羅的表現型を RDF 化して公開)、疾患・医薬品・環境物質データベース KEGG MEDICUS (研究者に加え医療従事者や一般の人々にも医薬品情報や疾患情報を提供、アクセス数は 20 万ユニーク IP/月以上)、が挙げられる。

3. 5. その他

NBDC では、発足以来一貫して公開のデータ等の共有を推進してきたが、一方でデータを直ちに広く公開することが困難な場合も存在する。データが非公開である段階から所在を把握しておくこと、および、公開前から共有に適したデータ整備を行っておくことが、将来のデータ公開と幅広い共有を促進するとの考えから、NBDC では「グループ内共有」の支援も 2016 (平成 28) 年度より開始している。これは、将来的な公開を前提に、公開までの間の、予め定めておいたメンバー間での共有を支援する取組みである。具体的には、日本医療研究開発機構 (AMED) の 9 つの連携分野の一つである「疾病克服に向けたゲノム医療実現化プロジェクト」、ならびに、JST 戦略的創造研究推進事業の CREST「環境変動に対する植物の頑健性の解明と応用に向けた基盤技術の創出」研究領域と連携した取組みを試行的に開始している。

このほか、NBDC 活動を広く周知するため、また、データベースの利活用を促進するため、学会や展示会等への出展、データベース講習会 (AJACS) の開催等を行っている。AJACS 等の講習会は、センター発足以来、約 40 回の開催実績がある。データ解析ニーズの大きい次世代シーケンサ (NGS) データについては、平成 26 年度より講習会を実施し、AJACS と共に人材支援にも貢献している。